IGCJ

Islamic Guidance and Counseling Journal
https://journal.iaimnumetrolampung.ac.id/index.php/igcj

Construct Validity Testing of the Theories of Intelligence Scale (TIS) on High School Students: Strategies for Addressing Misfit with the Rasch Model

Marifatin Indah Kholili*[1], Nandang Rusmana[1]
Ahman Ahman[1], Nandang Budiman[1]
Eka Sakti Yudha[1], Citra Tectona Suryawati[2]
[1] Universitas Pendidikan Indonesia
[2] Universitas Sebelas Maret, Indonesia
marifatin.ink23@upi.edu*

Article Information:
Received 2024-12-10
Revised 2025-06-20
Published 2025-09-07

Keywords:
Fixed Mindset, Growth Mindset, Rasch, Theories of Intelligence Scale

Abstract

The Theories of Intelligence Scale (TIS) has been widely used to measure mindsets in various countries; however, further research is still needed to examine its validity and reliability specifically within the Indonesian student population. TIS was chosen because it was directly developed by Carol Dweck, a leading figure in mindset theory, and is designed to clearly distinguish between fixed and growth mindsets. This study aims to test the validity and reliability of the Indonesian version of the TIS among secondary school students, while also presenting the stages of Rasch analysis and alternative options if the data do not fully meet Rasch model assumptions. A total of 681 students participated in this study, consisting of 271 males and 410 females, including 192 junior high school students, 243 senior high school students, and 246 vocational high school students. The Indonesian-translated TIS instrument consists of 8 items rated on a 6-point response scale. Data were analyzed using the Rasch Model to evaluate the rating scale structure, person fit, item fit, item measures, unidimensionality, Differential Item Functioning (DIF), and overall instrument reliability. The results indicate that the TIS generally meets the criteria of a sound measurement instrument. However, certain aspects require attention, particularly the need for item adjustments or further refinement for respondent groups with different characteristics. Future research with larger and more diverse samples is recommended to enhance the generalizability of the findings.

How to cite:	Kholili, M. I., Rusmana, N., Ahman, A., Budiman, N., Yudha, E. S., & Suryawati, C. T. (2025). Construct Validity Testing of the Theories of Intelligence Scale (TIS) on High School Students: Strategies for Addressing Misfit with the Rasch Model. *Islamic Guidance and Counseling Journal*, 8(2). https://doi.org/10.25217/0020258686400
E-ISSN:	2614-1566
Published by:	Institut Agama Islam Ma'arif NU (IAIMNU) Metro Lampung

INTRODUCTION

A mindset is a foundational belief regarding an individual's capability or ability to determine goals and behaviors. The right mindset serves as a crucial foundation as it fosters personal and professional development, encourages viewing challenges as learning opportunities, and enhances adaptability, which are essential for success in dynamic environments (Katsumata & Teixeira, 2024). The definition of mindset can be identified through expert opinions, each exploring the phenomenon from different perspectives and asserting distinct truths (Buchanan, 2024; Mappaenre et al., 2023). Initially, the term mindset combined the word "mind" with the psychological concept of "set" (McKeachie, 1990). Over time, it has become widely used in English. Mindset is defined as a 'mental attitude in the form of habits or characteristics that determine how an individual interprets and responds to situations (Noble, 2015). It is a conceptual tool useful for understanding how people perceive specific events and issues of practical concern (Fisher, 1988). Mindset is closely related to mental paradigms, emerging from prior experiences (Snyder, 2002), cultural beliefs (Fisher, 1988), and the formation of one's thought processes.

Understanding the formation of mindset requires examining its theoretical foundations. The theory of mindset is closely related to motivation theory (Dweck, 2006), particularly the concept of the malleability of abilities (Kapasi & Pei, 2021). Mindset theory represents a socio-cognitive approach, rooted in individual differences in beliefs and values influencing goals and behaviors. This theory, often referred to as implicit theory or self-theory, was introduced by Dweck (2000). According to self-theory, intelligence and other relatively stable traits are categorized as a fixed mindset / entity theory, while the belief that abilities can be developed through effort is termed a growth mindset / incremental theory (Macnamara & Rupani, 2017). Dweck's mindset theory centers on the idea that personal beliefs influence how individuals respond to challenging situations, often applied in academic contexts (Merrick, 2016).

Research on growth mindset has been conducted across diverse subjects and backgrounds, reflecting its broad applicability. Dweck highlights the implications of mindset in education, emphasizing that teaching and learning processes are more effective when both teachers and students adopt a growth mindset. She stresses the necessity for educators to foster a growth mindset to create an environment where students are encouraged to embrace challenges, accept constructive criticism, and persevere through failures (Dweck, 2015). A teacher's growth mindset positively correlates with mastery-oriented goal structures, suggesting that such teachers are more likely to adopt instructional practices emphasizing student learning and mastery (Bardach et al., 2024). Teachers' role in supporting a growth mindset significantly impacts students' mindset, engagement, and academic success (Setyorini & Khuriyah, 2023; Vestad & Bru, 2023). Feedback from teachers should prioritize the learning process and effort over outcomes (Azizah & Mardiana, 2024; Rhew et al., 2018). Teachers who emphasize learning processes and maintain high expectations for all students contribute to the development of students' growth mindsets (Haimovitz & Dweck, 2017).

Findings indicate that fostering a growth mindset regarding academic abilities can significantly influence students' perspectives on motivation and academic performance (Aditomo, 2015). Growth mindset interventions are vital for academic achievement, as they can teach students to adopt a growth-oriented perspective. Such a mindset encourages students to pursue learning goals, exert effort in overcoming challenges, and demonstrate resilience following failures (Macnamara & Burgoyne, 2023). Studies also explore how a growth mindset positively affects student behavior, fostering deeper learning, openness to feedback, and persistence in the face of setbacks (Campbell et al., 2019), leading to enhanced motivation (Rhew et al., 2018). This mindset increases students' resilience in confronting challenges (Yeager & Dweck, 2012). It plays important roles in educational settings, particularly for those who may struggle academically. A growth mindset impacts various aspects of student development, including academic motivation (Meyer & Stutts, 2024), self-efficacy (Mandeville et al., 2018), resilience (Yeager & Dweck, 2012), self-regulation (Astuti et al., 2024), academic performance (Canning et al., 2024; Hudig et al., 2022), persistence, and grit in solving problems (Larberg & Sherlin, 2021; YalcİN & Yilmaz, 2023). Additionally, it supports personal growth (Glerum et al., 2019) and adaptability (Mohamoud, 2024). Beyond academics, a growth mindset can positively affect other life areas, reducing stress (Meyer & Stutts, 2024; Sholihah et al., 2025), enhancing psychological well-being (Ali et al., 2023), and maintaining mental health (Blake, 2022; Lai et al., 2022). To identify whether individuals possess a growth mindset, appropriate instruments are required for accurate assessment.

Previous research has focused on developing instruments to assess growth mindset across various populations. Muradoglu et al. (2024) created the Growth Mindset Scale for Young Children (GM-C) for children aged 4–6 in the United States and South Africa, involving 220 participants. The GM-C scale includes four factors: beliefs about low instability of ability, low flexibility of ability, high instability of ability, and high flexibility of ability, with data analyzed using Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). Midkiff et al. (2018) examined Dweck's implicit theories of intelligence scale using an 8-item instrument with 6 response options, with a sample of 1,260 university students. Their analysis employed bifactor Item Response Theory and EFA. Another study (Chen et al., 2021) developed a growth mindset instrument for students and teachers in China, covering six dimensions: motivation, attitude, grit, challenge, adversity, and positive mindset. The study involved 654 participants, including 321 students and 266 teachers, with data analyzed through CFA and Structural Equation Modeling. In Norway, researchers tested the TIS with 723 participants aged 16–85, using Pearson correlation for analysis (Sigmundsson & Haga, 2024). Similarly, the TIS was examined in Indonesia with 213 participants aged 15–45 using CFA (Sigmundsson & Haga, 2024). Additionally, a growth mindset instrument was tested in Turkey using a sample of 1,145 students aged 14–22. This study applied both EFA and CFA for analysis (Yilmaz, 2022).

The use of the Rasch Model to analyze TIS data represents a novel approach that has not been previously employed. This approach enables a more detailed and objective analysis of each question item, including the detection of DIF (Differential Item Functioning). It is essential to ensure that the items in TIS are not biased against the characteristics of Javanese culture. Choosing TIS, combined with Rasch's analysis, makes a significant methodological contribution. The primary reason for choosing the TIS (Theories of Intelligence Scale) over other instruments, such as GM-C (Growth Mindset Scale for Young Children), is that the TIS is the original instrument developed by Carol Dweck, the primary pioneer of mindset theory. It makes it the most authentic and relevant tool for measuring the concept of mindset directly according to its original theoretical framework.

Instruments developed in different countries or regions must be validated to align with cultural contexts, ensuring accurate and reliable measurement. Cultural factors significantly influence the development of growth mindset (Buchanan, 2023; Zakrajsek, 2017) as intelligence and mindset can vary across cultural settings (Aditomo, 2015). This study aims to evaluate the applicability of TIS as a measurement tool for students in Indonesia. The study introduces a novel approach by emphasizing the research sample, geographic context, cultural specifications, and data analysis techniques. It focuses on testing the TIS with secondary school students in the ex-residency area of Surakarta, which comprises seven regencies in Central Java Province. The majority of respondents belong to the Javanese cultural group. This research will employ the Rasch model for data analysis, a method not previously utilized in validating the TIS.

METHODS

Instrument Translation Process

The instrument used was the Theories of Intelligence Scale (TIS) developed by Dweck (2000). This instrument assesses individuals' tendencies to view intelligence in two categories: as something that can develop (growth mindset) or as something fixed (fixed mindset). The instrument consisted of 8 reference items: (1) You have a certain amount of intelligence, and you can't really do much to change it; (2) Your intelligence is something about you that you can't change very much; (3) No matter who you are, you can significantly change your intelligence level; (4) To be honest, you can't really change how intelligent you are; (5) You can always substantially change how intelligent you are; (6) You can learn new things, but you can't really change your basic intelligence; (7) No matter how much intelligence you have, you can always change a quite bit; (8) You can change even your basic intelligence level considerably. Responses were collected using a 6-point Likert scale: strongly agree, agree, mostly agree, mostly disagree, strongly disagree. The instrument contained two dimensions: Growth mindset (items 3, 5, 7, and 8) and Fixed mindset (items 1, 2, 4, and 6).

The adaptation process began with obtaining permission and notifying Carol Dweck via email about the intention to translate the instrument into Indonesian. A professional translator then translated the instrument into Indonesian Language, followed by back-translation into English to ensure the terminology and intent aligned with the original instrument. To ensure readability, the translated instrument was tested with four students. Feedback was collected qualitatively to identify whether the translation was understandable or required adjustments. Suggestions included replacing the term "*Saudara*" with "*Anda*" in the statements for clarity and providing a definition of "*intelligence*" in the instructions. The revised instrument was then finalized and distributed to respondents.

Procedure and Participants

Participants were selected using random sampling, beginning with contacting school principals or guidance counselors in each school. The instrument was distributed online via Google Forms with assistance from the schools. Participants were provided with information about the researcher's identity, the purpose of data collection, and assurances regarding the confidentiality of the data they provided. Before completing the instrument, participants filled out personal information, including their name (optional), gender, class, ethnicity, school of origin, and qualitative questions related to mindset. The analysis involved 192 junior high school students, 243 senior high school students, and 246 vocational school students. Among the respondents, 271 were male, and 410 were female, with ages ranging from 12 to 18 years. The respondents came from seven districts within the ex-residency area of Surakarta in Central Java Province, Indonesia, totaling 681 students.

Data Analysis

The data in this study were analyzed using the Rasch model. It analyzes psychometric techniques and how these techniques can be used by researchers. The Rasch model is specifically used to document and evaluate the measurement properties of instruments. This analysis is also commonly known as the Rasch Model or Rasch Measurement, developed by George Rasch, a Danish mathematician, in 1960 (Bond & Fox, 2015). Built upon Item Response Theory (IRT), it examines the relationship between item characteristics and respondent abilities (Rangka et al., 2023). While Rasch analysis is often categorized under the umbrella of IRT models, it differs philosophically from the typical 1-parameter, 2-parameter, or 3-parameter IRT models. Unlike IRT models, which are adjusted to fit the data by adding parameters, the Rasch model defines measurement without modification to align with the data, making it a strict definition of measurement (Boone, 2016). Two common parameterizations of the Rasch model, the partial credit model and the rating scale model, are identical at the level of analyzing an individual's response to a single item (Engelhard & Wind, 2017). The analysis is grounded in two key theorems: the individual's ability or agreement level and the difficulty level of the item (Cavanagh & Waugh, 2011).

Rasch analysis provides comprehensive insights into the instrument, including model fit, reliability, scale bias, standard deviation, unidimensionality, and rating scale analysis, by assessing both item and respondent measurements (Boone et al., 2014). Additionally, Rasch analysis identifies misfitting items, evaluates their alignment with expected outcomes, and reveals respondent misunderstandings (Aminah et al., 2024; Sumintono & Widhiarso, 2014). The analysis for this study was conducted using Winsteps software, developed by Mike Linacre (Linacre, 2012).

RESULTS AND DISCUSSION

RESULTS

Rating Scale

The initial step in the Rasch analysis involved assessing the functioning of the available response categories by examining whether each option was used consistently and in the expected order. The adapted instrument used a six-point Likert scale and was completed by 1,013 secondary school students during the initial data collection. The result of rating scale are presented in table 1.

The rating scale was analyzed by observing the Andrich Threshold, which represents the difficulty level required for a respondent to transition from one response category to the next. According to the Polychotomous Threshold Formulation, the threshold values should begin with "None" and systematically progress from negative to positive (Andrich, 2025). The analysis, as presented in Table 1, revealed a threshold disorder between response categories 4 and 5. The threshold decreased from 0.54 to 0.45, indicating a disruption in the expected increasing order of category thresholds. This may suggest that respondents found it difficult to differentiate between these adjacent categories, possibly leading to overlapping usage and reduced category clarity.

Person Fit

The next step ensured that respondents fit the Rasch model. Understanding the statement items and response options can be influenced by factors such as an excessive number of options, confusion in selecting responses, or lack of effort in completing the instrument. Another technique to evaluate the instrument's quality is Person Fit, which identifies respondents displaying unusual response patterns. For instance, a pattern may indicate that a student concentrated for some time and then lost focus, or that they were guessing answers instead of responding sincerely (Boone, 2016). The Person Fit analysis used the Outfit Mean Square (MNSQ) criterion, with acceptable values between 0.5 and 1.5 (Sumintono, 2016).

Based on the Person Fit table output from Winsteps, 332 respondents were identified as outliers. This process was corroborated by the Scalogram analysis, which examined

Table 1. Rating Scale Analysis (N=1013)

Category Label	Score	Observed Count	Obsvd Avrge	Sample Expect	Infit Mnsq	Outfit Mnsq	Andrich Thresholid	Category Measure
1	1	1212 15	-1.22	-1.32	1.17	1.10	None	(-3.08)
2	2	2655 33	-.81	-.74	.86	.91	-1.84	-1.26
3	3	2079 26	-.39	-.37	.90	.91	-.31	-.18
4	4	1003 12	.06	-.02	.87	.87	.54	.49
5	5	756 9	.43	.34	.84	.85	.45	1.23
6	6	399 5	.57	.71	1.23	1.40	1.17	(2.56)

Table 2. Rating Scale Analysis (N=681)

Category Label	Score	Observed Count	Obsvd Avrge	Sample Expect	Infit Mnsq	Outfit Mnsq	Andrich Thresholid	Category Measure
1	1	775 14	-1.32	-1.41	1.10	1.06	None	(-3.19)
2	2	1700 31	-.84	-.80	.94	.98	-1.95	-1.36
3	3	1453 27	-.45	-.39	1.01	1.03	-.44	-.23
4	4	731 13	.07	.00	.94	.93	.49	.51
5	5	538 10	.48	.41	.87	.81	.51	1.32
6	6	251 5	.79	.85	1.04	1.00	1.39	(2.75)

respondents' tendencies in answering items. After removing the outlier respondents, 681 participants remained for further analysis. This selection process significantly impacted the Rasch analysis results. The revised Andrich Threshold values began with "None" and systematically progressed from negative to positive, indicating that the analysis could proceed to the next stage. Rating scale analysis with 681 respondents presented on table 2.

Item Fit and Item Measure

The subsequent step involved ensuring item acceptability by verifying that each item aligned with the Rasch model. This analysis considered three psychometric attributes: Outfit Mean Square (MNSQ) (0.5 < MNSQ < 1.5), Outfit Z-Standard (ZSTD) (-2.0 < ZSTD < +2.0), and Point Measure Correlation (Pt Measure Corr) (0.32 < Pt Measure Corr < 0.8). Items falling outside these thresholds were categorized as misfitting and required revision and revalidation (Rangka et al., 2023; Sumintono, 2018).

Based on the item fit analysis on Table 3, the mean score was 1956.8, indicating the average total score obtained by participants, which serves as a reference point for comparing individual and group performance. Mean Measure: A value of 0.00 indicated that the average latent ability of participants was at the midpoint of the scale, suggesting that participant abilities were evenly distributed around this midpoint. P.Std: A value of 0.56 indicated significant variation in participant abilities, meaning some participants performed well above the average while others performed well below it. The analysis results showed that all eight items of the Theories of Intelligence Scale (TIS) met the criteria for Outfit MNSQ and Pt Measure Correlation. For Outfit ZSTD, acceptable values ranged from -2.0 to +2.0 for samples of 30 to 300 respondents (Bond & Fox, 2015). However, for samples exceeding 300 respondents, ZSTD could be excluded from consideration. The next stage of analysis focused on item measurement presented on table 4.

In Rasch analysis, item measure indicates the difficulty level of an item. Higher (more positive) item measure values denote more difficult items, while lower (more negative) item measure values indicate easier ones. In this item measure analysis, Item 4 had the lowest measure value (-1.16), making it the easiest item among all. Most participants were likely able to respond to this item correctly. Conversely, Item 5 and Item 3 had the highest measure values (0.64 and 0.57), indicating they were the most challenging items. Only participants with very high abilities were likely to answer these items correctly.

The remaining items (Item 7: 0.39; Item 8: 0.18; Item 1: -0.01; Item 6: -0.13; Item 2: -0.49) exhibited a range of intermediate difficulty levels. The findings revealed that the items in the Theories of Intelligence Scale (TIS) exhibited a broad range of difficulty levels. This variation provides useful insights for refining or enhancing the test instrument. Items deemed too easy or too difficult can be replaced or modified to achieve a more balanced difficulty distribution across the scale.

Unidimensionality

The next step involved evaluating the instrument's residual standard to assess unidimensionality. Unidimensionality signifies that all items in the questionnaire measure only

Table 3. Item Fit Order (N=681)

Entry Number	Total Score	Total Count	Jmle Measure	Model S.E	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	PtMeasure-Al CORR.	PtMeasure-Al EXP.	Exact Match OBS	Exact Match EXP	Item
5	1529	681	.64	.04	1.18	2.82	1.14	2.29	A .37	.45	40.2	45.0	i5
4	2949	681	-1.16	.04	1.14	2.78	1.14	2.68	B .55	.61	32.6	32.7	i4
2	2300	681	-.49	.04	1.13	2.60	1.11	2.18	C.60	.55	25.9	33.1	i2
3	1566	681	.57	.04	.95	-.74	.96	-.72	D .43	.45	42.4	44.1	i3
6	2016	681	-.13	.04	.94	-1.23	.96	-.74	d .52	.52	37.4	36.3	i6
1	1928	681	-.01	.04	.85	-2.84	.85	-2.81	c .56	.51	39.1	38.0	i1
7	1667	681	.39	.04	.81	-3.52	.79	-3.81	b. 50	.47	45.2	42.1	i7
8	1801	681	.18	.04	.81	-3.53	.81	-3.44	a .50	.49	42.9	40.1	i8
Mean	1956.8	681.0	.00	.04	.98	-.5	.97	-.5			38.2	38.9	
P.SD	412.4	.0	.56	.00	.14	2.6	.14	2.5			5.9	4.4	

Table 4. Item Measure (N=681)

Entry Number	Total Score	Total Count	Jmle Measure	Model S.E	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Ptmeasure-Al CORR.	Ptmeasure-Al EXP.	Exact Match OBS	Exact Match EXP	Item
5	1529	681	.64	.04	1.18	2.02	1.14	.229	.37	.45	40.2	45.0	15
3	1566	681	.57	.04	.95	-.74	.96	-.72	.43	.45	42.4	44.1	i3
7	1667	681	.39	.04	.81	-3.52	.79	-3.81	.50	.47	45.2	42.1	i7
8	1801	681	.18	.04	.81	-3.53	.81	-3.44	.50	.49	42.9	40.1	i8
1	1927	681	-.01	.04	.85	-2.84	.85	-2.81	.56	.51	39.1	38.0	i1
6	2016	681	-.13	.04	.94	-1.23	.96	-.74	.52	.52	37.4	36.3	i6
2	2300	681	-.49	.04	1.13	2.60	1.11	2.18	.60	.55	25.9	33.1	i2
4	2848	681	-1.16	.04	1.14	2.78	1.14	2.68	.55	.61	32.6	32.7	i4
Mean	1956.8	681.0	.00	.04	.98	-.5	.97	-.5			38.2	38.9	
P.SD	412.4	.0	.56	.00	.14	2.6	.14	2.5			5.9	4.4	

one latent construct under investigation. Additionally, Local Dependency requires that responses to specific items are independent of responses to other items within the questionnaire (Eugenio & Silvia, 2007). Unidimensionality is satisfied if Raw variance explained by measures exceeds 40% of the total explained variance (Aminah et al., 2024). Raw unexplained variance in the 1st to 5th contrast is ≤ 15% (Fisher, 2015).

Raw variance explained by measures: Reflects the variability explained by differences in participants' abilities. Raw variance explained by persons: Captures the variability attributed to individual participant differences. Raw variance explained by items: Indicates the variability explained by differences in item difficulty. Raw unexplained variance (total) represents the residual variability not accounted for by the model. A higher value indicates more "noise" or interference in the data that the model fails to capture. Unexplained variance in the 1st to 5th contrast: Provides details on residual variability partitioned across contrasts between response categories. High values in specific contrasts suggest potential issues with the model related to those response categories. Standardized residual presented on tabel 5.

The results of the unidimensionality analysis on table 5 indicate that the raw variance explained by measures was 45.2% (Eigenvalue: 6.58), The raw variance explained by persons was 9.3% (Eigenvalue: 1.35), while the raw variance explained by items was 35.9% (Eigenvalue: 5.23). For unexplained variance, the 1st contrast was 18.2% (Eigenvalue: 2.65), exceeding the acceptable threshold of 15%, indicating a lack of fit. In contrast, the unexplained variance in the 2nd to 5th contrasts was within the acceptable range, with values of 8.2%

Table 5. Standardized Residual

	Eigenvalue	Observed		Expected
Total raw variance in observations	14.5892	100.0%		100.0%
Raw variance explained by measures	6.5892	45.2%		44.2%
Raw variance explained by persons	1.3553	9.3%		9.1%
Raw variance explained by items	5.2339	35.9%		35.1%
Raw unexplained variance (total)	8.0000	54.8%	100.0%	55.8%
Unexplned variance in 1st contrast	2.6550	18.2%	33.2%	
Unexplned variance in 2nd contrast	1.1983	8.2%	15.0%	
Unexplned variance in 3rd contrast	1.0849	7.4%	13.6%	
Unexplned variance in 4th contrast	1.0263	7.0%	12.8%	
Unexplned variance in 5th contrast	0.7516	5.2%	9.4%	

Table 6. DIF Analysis by Grades

Person Classes	Summary Dif Chi-Squared	D.F	Prob	Between-Class/Group Item UNWTD MNSQ	ZSTD	Number	Name
3	.0000	2	1.0000	.0422	-1.62	1	I1
3	14.0012	2	.0009	7.1264	3.11	2	I2
3	3.0915	2	.2101	1.5712	.82	3	I3
3	4.4525	2	.1060	2.2695	1.28	4	I4
3	1.6756	2	.4291	.8471	.17	5	I5
3	3.0317	2	.2165	1.5262	.79	6	I6
3	3.7556	2	.1504	1.9088	1.05	7	I7
3	3.5621	2	.1658	1.8097	.99	8	I8

(Eigenvalue: 1.19), 7.4% (Eigenvalue: 1.08), 7% (Eigenvalue: 1.02), and 5.2% (Eigenvalue: 0.75), respectively. Based on these results, the raw variance explained by measures and the unexplained variance in the 2nd to 5th contrasts met the required standards. However, the unexplained variance in the 1st contrast did not meet the threshold.

DIF Analysis

The Differential Item Functioning (DIF) analysis was conducted to ensure the absence of bias in the instrument. This test examines item difficulty based on respondents' demographic profiles, providing valuable insights for mapping overall abilities by respondent characteristics. An item is considered to exhibit DIF (bias) if its p-value is less than 0.05 (Sumintono, 2018). The DIF analysis in this study focused on respondent characteristics based on gender (male and female) and educational level (junior high school, senior high school, and vocational high school). DIF Analysis presented on table 6.

Based on Table 6, Item 2 demonstrated a significant DIF, with a p-value of 0.0009, indicating a notable difference in item functioning among the three groups. Additionally, the large UNWTD MNSQ value (7.1264) for this item suggests severe misfit or unpredictable responding, especially for the group impacted by the DIF. This unusually high MNSQ value indicates that the observed responses for Item 2 deviate substantially from what is expected by the Rasch model for certain individuals or groups. The presence of DIF in an item can impact

Table 7. DIF Analysis by Grades Spesification

Person	Observations		Basaline		DIF	DIF	Infit		Outfit		Item	
Class	Count	Avarage	Expect	Measure	Size	S.E	MNSQ	ZSTD	MNSQ	ZSTD	Number	Name
A	191	1.75	1.77	-.01	0.00	.07	1.01	.16	1.01	.14	1	I1
B	239	2.05	2.03	-.01	0.00	.06	.70	-3.79	.71	-3.67	1	I1
C	242	1.74	1.75	-.01	0.00	.06	.90	-1.14	.87	-1.41	1	I1
A	191	2.23	2.30	-.49	.07	.07	1.31	-3.04	1.25	2.42	2	I2
B	239	2.84	2.63	-.49	-.17	.06	1.02	.23	1.01	.12	2	I2
C	242	2.13	2.28	-.49	-.13	.06	1.12	1.37	1.11	1.26	2	I2
A	191	1.25	1.26	.57	.00	.08	1.05	.47	1.03	.30	3	I3
B	239	1.38	1.45	.57	.08	.07	.88	-11.23	.87	-1.40	3	I3
C	242	1.31	1.24	.57	-.09	.07	.96	-.33	.99	-.11	3	I3
A	191	3.25	3.11	-1.16	-.12	.07	1.27	2.72	1.28	2.72	4	I4
B	239	3.38	3.46	-1.16	.06	.06	.99	-0.08	.99	-.11	4	I4
C	242	3.05	3.09	-1.16	.03	.06	1.19	2.21	1.19	2.12	4	I4
A	191	1.14	1.20	.64	.09	.09	1.29	2.33	1.24	1.94	5	I5
B	239	1.43	1.39	.64	-.05	.07	1.15	1.47	1.12	1.23	5	I5
C	242	1.19	1.18	.64	.00	.07	1.13	1.22	1.09	.90	5	I5
A	191	1.94	1.90	-.13	-.04	.07	1.02	.18	1.02	.26	6	I6
B	239	2.08	1.18	-.13	.08	.06	.92	-0.99	.93	-.81	6	I6
C	242	1.93	1.87	-.13	-.06	.06	.89	-1.19	.94	-.65	6	I6
A	191	1.50	1.40	.39	-.11	.08	.93	-.56	.86	-1.26	7	I7
B	239	1.53	1.61	.39	.09	.07	.76	-2.72	.78	-2.40	7	I7
C	242	1.38	1.38	.39	.00	.07	.76	-2.64	.74	-2.83	7	I7
A	191	1.48	1.59	.18	.13	.08	.94	-.53	.94	-.55	8	I8
B	239	1.88	1.83	.18	-.05	.06	.76	-2.83	.75	-2.96	8	I8
C	242	1.61	1.57	.18	-.04	.07	.77	-2.56	.79	2.37	8	I8

the validity and reliability of the test by introducing measurement errors in assessing the true abilities of specific groups.

Based on Table 7, Differential Item Functioning (DIF) was assessed using statistical significance (ZSTD value outside ±1.96, corresponding to p<.05) and practical significance (DIF measure greater than 0.5 logits). The analysis identified several items exhibiting DIF among the three groups (Group A: Junior High School, Group B: Senior High School, Group C: Vocational High School). Item 2 showed statistically significant negative DIF for Group A compared to both Group B and Group C (DIF = −0.49, ZSTD = −2.63 for both comparisons), indicating it was significantly more difficult for Group A. However, its DIF magnitude was below the practical significance threshold. Item 6 also demonstrated statistically significant DIF. It was more difficult for Group A than Group B (DIF = −0.13, ZSTD = −3.25) and easier for Group C than Group A (DIF = 0.11, ZSTD = 2.75). Similar to Item 2, its DIF magnitudes were below the practical significance threshold. Item 5 did not reach statistical significance in any comparison (ZSTD ranging from 1.47 to −1.53). However, it showed practically significant DIF, being substantially easier for Group A compared to Group B (DIF = 0.64) and substantially more difficult for Group C compared to Group A (DIF = −0.59). Item 7 (DIF = −0.76, ZSTD = −2.72) was significantly and practically more difficult for Group C than for Group B. Item 8 (DIF = −0.76, ZSTD = −2.83) was significantly and practically more difficult for Group B than for Group C. No other items displayed statistically significant DIF across the groups. The presence of DIF in these items highlights potential measurement bias, impacting the validity and reliability of the test in assessing the true abilities of specific groups. To visually complement these findings, Figure 1. illustrates the DIF patterns across the three groups.

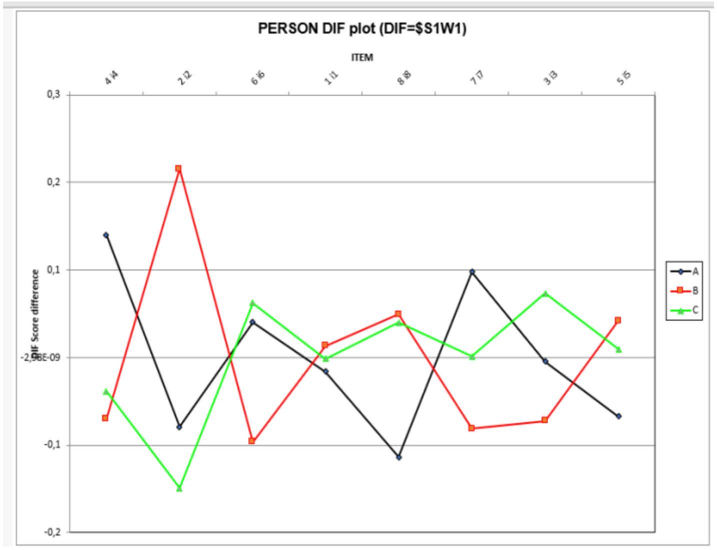

Figure 1. Person DIF Plot

Summary Statistic

The Cronbach's Alpha value (KR-20) of 0.67 indicates that the test reliability or the overall person-item interaction is fair (notes: < 0.5: poor; 0.5-0.6: weak; 0.6-0.7: fair; 0.7-0.8: good, > 0.8: excellent). With Person reliability at 0.67 and Item reliability at 0.99, based on the Fisher table, the consistency of respondent answers falls into the fair category, and the item reliability is in the outstanding category (notes: < 0.67: weak; 0.67-0.88: fair; 0.8-0.90: good, 0.91-0.94: excellent; > 0.94: outstanding) (Fisher, 1992; Kholili et al., 2023). The raw variance by measure is 45.2%, which is above the 40% threshold, indicating that the instrument is suitable for measuring what it is intended to measure. The summary statistic of Theories of Intelligence Scale (TIS) presented on table 7.

Table 7. Summary Statistic (N= 681, i=8)

	Reliabillity	Separation Index	Mean Measure	Cronbach's Alpha	Raw Explained by measures
Person	0.67	1.42	-0.46	0.67	45.2
Item	0.99	13.93	0.00		

Based on table 7, The item separation index is 13.93, and the person separation index is 1.42. These indices indicate the extent to which the instrument differentiates between individual latent abilities and how the distribution of items separates the abilities of each respondent. The average ability of the respondents is -0.46 logits, suggesting that respondents generally have slightly below-average ability to answer the items correctly.

DISCUSSION

The aim of this study was to examine the validity and reliability of the TIS using the Rasch Model analysis. In preparing for research, researchers often face various challenges in data collection, statistical calculations, and obtaining accurate results. At times, they may encounter different errors (Mohajan, 2017). Similarly, the process of instrument validation may not always proceed as expected. Instead of solely presenting ideal fit data, this article addresses these common challenges by demonstrating practical solutions encountered during the Rasch model analysis, such as managing misfit respondents (e.g., through outlier removal) and evaluating the functioning of response categories. These approaches can serve as alternative strategies for other researchers facing similar issues in their instrument development and validation processes

The use of an instrument to measure a particular variable heavily depends on the validity and reliability results of the instrument (Amir-Behghadami et al., 2020). Key indicators of the quality of a measurement tool are its precision in reliability and validity (Mohajan, 2017). The Rasch model analysis is one of the analytical models that can be used to assess the quality of an instrument (Fischer & Molenaar, 1995). Interpretation and analysis in Rasch must be an iterative process until all fit criteria are fully met. During the analysis process, data that does not meet the ideal criteria may arise. Possible misfit can be influenced by various factors, such as answer choices, the number of items, and respondent characteristics.

Answer choices for psychological scales often use Likert scales with four or five response options. The TIS instrument developed by Carol Dweck includes six answer choices. Having too many answer options can lead to confusion and difficulty for respondents in selecting the most appropriate answer (McClelland, 1973). Studies examining response scales ranging from 2 to 11 options show that scales with 2 to 5 options tend to exhibit a decline in psychometric precision, while no advantage is found for scales with more than 6 options (Simms et al., 2019). Research on answer choices suggests multiple-choice questions (MCQ) with three optimal options maintain similar psychometric properties compared to four or five options, while improving test efficiency and reducing preparation time for distractors (Baghaei & Amrahi, 2011; Vyas & Supe, 2008). Other studies recommend four answer options as the optimal number to balance guessing and reliability. Probabilistic analysis indicates that four options minimize the likelihood of respondents guessing randomly (Zhao, 2006). In line with these studies, researchers developing measurement tools are recommended to limit answer options to a maximum of five on the rating scale.

In the analysis of the TIS instrument, prior to respondent selection, the Andrich threshold for answer choice 5 (disagree) showed a decrease in score compared to answer choice 4, indicating inconsistency in respondent responses. This suggests that respondents may not provide consistent answers according to the scale used (Das, 2001). This could be due to several factors: respondents may not fully understand the intent of the question, or they may make errors in completing the questionnaire. To address this, researchers could consider piloting the questionnaire with a small group to identify confusing or ambiguous questions, detect options with unusual Andrich thresholds, and consider revising or removing them. It would also be important to ensure that respondents have sufficient motivation to answer the questionnaire seriously.

Instrument validity can be compromised when respondents are unwilling or unable to provide adequate responses to the instrument's questions. Response validity has become a growing concern in empirical research (Edwards, 2019). Respondent tendencies such as carelessness, guessing, multidimensionality, and others lead to specific deficiencies in the invariant hypothesis for the appropriate subdivisions of the total test (Fischer & Molenaar, 1995). If the characteristics of the respondents do not accurately reflect their true state, it can impact the instrument's validity, affect the reliability of the data collected, and influence conclusions drawn (Mokkink et al., 2021). Respondent bias in surveys is highly likely. Social bias in surveys refers to the tendency of respondents to provide answers that align with social norms rather than reflect their true opinions or behaviors. This bias is influenced by various factors, including question design, data collection methods, and individual respondent characteristics (Gittelman et al., 2015). This indicates that respondents may not provide answers fully consistent with the intended scale (Wu & Tang, 2016).

This study not only contributes to the theoretical understanding of Rasch Model analysis but also has significant practical implications for the development and use of the TIS instrument. The findings from this research can serve as a guide for researchers in developing better instruments and producing more accurate findings. Furthermore, this study emphasizes the importance of ensuring data quality in Rasch Model analysis to maintain the validity and reliability of the results.

CONCLUSION

Based on the results of the Rasch Model analysis, the TIS instrument generally meets the criteria of a good instrument. However, several aspects need attention: adjustments should be made to the number of response options to avoid confusion for respondents, item development or modification should be considered for respondent groups with different characteristics, and further research with larger and more diverse samples is needed to confirm the analysis results and enhance the generalization of findings. By making the necessary improvements and developments, this instrument can become a more effective tool for measuring intelligence theory across different populations. The application of the Rasch Model analysis is a powerful tool for testing the validity and reliability of instruments, but any challenges encountered in its application can be addressed by following the recommendations provided. This study demonstrates that issues such as too many response options, inconsistency in respondents' answers, and respondent bias can affect the quality of data and analysis results. Therefore, researchers need to carefully consider instrument design, the choice of response scale, and respondent characteristics to minimize bias and improve result accuracy. This study provides several recommendations for future researchers, such as considering a response scale with only four to five options, conducting careful pre-testing of the instrument, and ensuring respondent motivation. The instrument distribution and completion process should also be more closely monitored, and researchers need to be mindful of potential respondent bias and explore ways to minimize it.

ACKNOWLEDGEMENT

The authors would like to express gratitude for the financial support that made this research possible. This research was fully funded by The Indonesian Education Scholarship (BPI), Center for Higher Education Funding and Assessment (PPAPT), The Indonesian

Endowment Fund for Education (LPDP), Ministry of Higher Education, Science and Technology of the Republic of Indonesia.

AUTHOR CONTRIBUTION STATEMENT

MIK plays the role of the principal researcher, designing the research, CTS collects data and analyzes the results. MIK, ESY made a significant contribution to data interpretation and the writing of discussion sections. NR, A, and NB contributed to the literature review and revision of the manuscript. All authors discuss the results and contribute to the final manuscript.

REFERENCES

Aditomo, A. (2015). Students' Response to Academic Setback: 'Growth Mindset' as a Buffer Against Demotivation. *International Journal of Educational Psychology*, *4*(2), 198-222. https://doi.org/10.17583/ijep.2015.1482

Ali, S., Zarar, N., & Siddiqui, G. K. (2023). Impact of Growth Mind-Set on University Students' Psychological Well-Being. *Annals of Human and Social Sciences*, *4*(III). https://doi.org/10.35484/ahss.2023(4-III)19

Aminah, S., Hidayah, N., Hanurawan, F., & Indreswari, H. (2024). Tailoring of the career adaptabilities scale for Indonesian youth. *Children and Youth Services Review*, *166*(January), 107914. https://doi.org/10.1016/j.childyouth.2024.107914

Amir-Behghadami, M., Gholizadeh, M., & Janati, A. (2020). Methodological Issues on the Importance of Instrument Validation in Cross-Sectional Health Research. *Bull Emerg Trauma*, *8*(1), 49-50. https://doi.org/10.29252/beat-080109

Andrich, D. (2025). A Rating Formulation for Ordered Response Categories. *Psychometrika*, *43*(4), 561-573. https://doi.org/10.1007/bf02293814

Astuti, F. W., Sunawan, & Mugiarso, H. (2024). View of Effectiveness of Group Guidance on Growth Mindset for Improving Students' Effort Regulation. *Jurnal Bimbingan Konseling*, *13*(1), 87–94. Google Scholar

Azizah, I. a., & Mardiana, D. (2024). Learning Transformation: Increasing Student Achievement through Discovery Learning. *Dirasah International Journal of Islamic Studies*, *2*(2), 155-166. https://doi.org/10.59373/drs.v2i2.42

Baghaei, P., & Amrahi, N. (2011). The effects of the number of options on the psychometric characteristics of multiple choice items. *Psychological Test and Assessment Modeling*, *53*(2), 192–211. Google Scholar

Bardach, L., Bostwick, K. C. P., Fütterer, T., Kopatz, M., Hobbi, D. M., Klassen, R. M., & Pietschnig, J. (2024). A Meta-Analysis on Teachers' Growth Mindset. *Educational psychology review*, *36*(3). https://doi.org/10.1007/s10648-024-09925-7

Blake, T. G. (2022). Growth Mindset and Positive Mental Health. *Sustainability (Switzerland*, *11*(1), 1–14. Google Scholar

Bond, T. G., & Fox, C. M. (2015). Applying the Rasch model: fundamental measurement in the human sciences, Third Edition. *Maternal & Child Nutrition*, *11*(ue supplement S2)). https://doi.org/10.4324/9781410614575

Boone, W. J. (2016). Rasch Analysis for Instrument Development: Why, When, and How? *CBE Life Sci Educ*, *15*(4). https://doi.org/10.1187/cbe.16-04-0148

Boone, W. J., Yale, M. S., & Staver, J. R. (2014). Rasch Analysis in the Human Sciences. https://doi.org/10.1007/978-94-007-6857-4

Buchanan, A. (2023). The History of Mindset: A Critical Review. *Middle East Journal of Positive Psychology*, *9*, 79–99. Google Scholar

Buchanan, A. (2024). *What is mindset? 100 definitions from the field*. In Handbook of Mindset Research. https://osf.io/preprints/psyarxiv/5xeqv

Campbell, A., Craig, T., & Collier-Reed, B. (2019). A framework for using learning theories to inform 'growth mindset' activities. *International Journal of Mathematical Education in Science and Technology*, *51*(1), 26-43. https://doi.org/10.1080/0020739x.2018.1562118

Canning, E. A., White, M., & Davis, W. B. (2024). Growth Mindset Messages from Instructors Improve Academic Performance Among First-Generation College Students. *CBE Life Sci Educ*, *23*(2), ar14. https://doi.org/10.1187/cbe.23-07-0131

Cavanagh, R. F., & Waugh, F. (2011). Applications of Rasch Measurement in Learning Environments Research. Google Scholar

Chen, S., Ding, Y., & Liu, X. (2021). Development of the growth mindset scale: evidence of structural validity, measurement model, direct and indirect effects in Chinese samples. *Current Psychology*, *42*(3), 1712-1726. https://doi.org/10.1007/s12144-021-01532-x

Das, S. (2001). Inconsistency in Response When Questions are Repeated with the Number of Options Changing. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2161160

Dweck, C. S. (2000). *Self-Theories: Their Role in Motivation, Personality, and Development*. Psychology Press -Taylor & Francis Group. https://doi.org/10.4324/9781315783048

Dweck, C. S. (2006). *Mindset: The New Psychology of Success*. Random House. Google Scholar

Dweck, C. S. (2015). Carol Dweck Revisits the "Growth Mindset. *Education Week*, *35*(ue 05)). Google Scholar

Edwards, J. R. (2019). Response invalidity in empirical research: Causes, detection, and remedies. *Journal of Operations Management*, *65*(1), 62-76. https://doi.org/10.1016/j.jom.2018.12.002

Engelhard, G., & Wind, S. A. (2017). Invariant Measurement with Raters and Rating Scales. In *Invariant Measurement with Raters and Rating Scales* (pp. 327-336). https://doi.org/10.4324/9781315766829-14

Eugenio, B., & Silvia, G. (2007). Unidimensionality in the Rasch Model: How to Detect and Interpret. *STATISTICA*, *LXVII*(3). https://doi.org/10.6092/issn.1973-2201/3508

Fischer, G. H., & Molenaar, I. W. (1995). Rasch Models. In *Rasch Models*. https://doi.org/10.1007/978-1-4612-4230-7

Fisher, G. (1988). *Mindsets : The Role of Culture and Perception in International Relations* (2nd ed.). Intercultural Press. Google Scholar

Fisher, W. P. (1992). *Rasch Measurement Transactions*, *6*(3), 228.

Fisher, W. P. (2015). A Probabilistic Model of the Law of Supply and Demand. *Transactions of the Rasch Measuremen*, *29*(1), 1499–1514.

Gittelman, S., Lange, V., Cook, W. A., Frede, S. M., Lavrakas, P. J., Pierce, C., Thomas, R. K., Terhanian, G., Bacon, C., & Fulgoni, G. (2015). Accounting for Social-Desirability Bias in Survey Sampling. *Journal of Advertising Research*, *55*(3), 242-254. https://doi.org/10.2501/jar-2015-006

Glerum, J., Loyens, S. M. M., & Rikers, R. M. J. P. (2019). Mind your mindset. An empirical study of mindset in secondary vocational education and training. *Educational Studies*, *46*(3), 273-281. https://doi.org/10.1080/03055698.2019.1573658

Haimovitz, K., & Dweck, C. S. (2017). The Origins of Children's Growth and Fixed Mindsets: New Research and a New Proposal. *Child Dev*, *88*(6), 1849-1859. https://doi.org/10.1111/cdev.12955

Hudig, J., Scheepers, A. W. A., Schippers, M. C., & Smeets, G. (2022). Motivational mindsets, mindset churn and academic performance: The role of a goal-setting intervention and purpose in life. *Current Psychology*, *42*(27), 23349-23368. https://doi.org/10.1007/s12144-022-03462-8

Kapasi, A., & Pei, J. (2021). Mindset Theory and School Psychology. *Canadian Journal of School Psychology*, *37*(1), 57-74. https://doi.org/10.1177/08295735211053961

Katsumata, R. N., & Teixeira, S. L. (2024). The Importance of Developing Growth Mindset. *Editora*, 284–294. https://doi.org/10.56238/sevened2024.029-016

Kholili, M. I., Asrowi, & Susilo, A. T. (2023). Breaking Ground in Skill Assessment : The Construction and Validation of a Problem- Solving Scale for University Students. *Konselor*, *12*(4). https://doi.org/10.24036/0202312452-0-86

Lai, X., Nie, C., Huang, S., Li, Y., Xin, T., Zhang, C., & Wang, Y. (2022). Effect of Growth Mindset on Mental Health Two Years Later: The Role of Smartphone Use. *Int J Environ Res Public Health*, *19*(6). https://doi.org/10.3390/ijerph19063355

Larberg, J. L., & Sherlin, L. H. (2021). Grit and Growth Mindset Contribution to School Counseling Services. *SAGE Open*, *11*(2). https://doi.org/10.1177/21582440211014512

Linacre, J. M. (2012). *Winsteps (Version 3.74*. Software. Google Scholar

Macnamara, B. N., & Burgoyne, A. P. (2023). Do growth mindset interventions impact students' academic achievement? A systematic review and meta-analysis with recommendations for best practices. *Psychol Bull*, *149*(3-4), 133-173. https://doi.org/10.1037/bul0000352

Macnamara, B. N., & Rupani, N. S. (2017). The relationship between intelligence and mindset. *Intelligence*, *64*(July), 52-59. https://doi.org/10.1016/j.intell.2017.07.003

Mandeville, D., Perks, L., Benes, S., & Poloskey, L. (2018). The Mindset and Intellectual Development Scale (MINDS): Metacognitive Assessment for Undergraduate Students. *International Journal of Teaching and Learning in Higher Education*, *30*(3), 497–505. Google Scholar

Mappaenre, A., Hasanah, A., Samsul Arifin, B., Nuraini, Y., & Satria Wiwaha, R. (2023). The Implementation of Character Education in Madrasah. *Attadrib: Jurnal Pendidikan Guru Madrasah Ibtidaiyah*, *5*(2), 166-181. https://doi.org/10.54069/attadrib.v5i2.302

McClelland, D. C. (1973). Testing for competence rather than for "intelligence". *Am Psychol*, *28*(1), 1-14. https://doi.org/10.1037/h0034092

McKeachie, W. J. (1990). Learning, Thinking, and Thorndike. *Educational Psychologist*, *25*(2), 127-141. https://doi.org/10.1207/s15326985ep2502_2

Merrick, J. S. (2016). Mindset Matters: Measuring Anxiety Mindsets. Google Scholar

Meyer, H. H., & Stutts, L. A. (2024). The Effect of Mindset Interventions on Stress and Academic Motivation in College Students. *Innovative Higher Education*, *49*(4), 783-798. https://doi.org/10.1007/s10755-024-09706-8

Midkiff, B., Langer, M., Demetriou, C., & Panter, A. T. (2018, 2018). *An IRT analysis of the growth mindset scale* Springer Proceedings in Mathematics and Statistics, http://dx.doi.org/10.1007/978-3-319-77249-3_14

Mohajan, H. K. (2017). Two Criteria for Good Measurements in Research: Validity and Reliability. *Annals of Spiru Haret University. Economic Series*, *17*(4), 59-82. https://doi.org/10.26458/1746

Mohamoud, A. M. (2024). The Impact of Growth Mindset Interventions on Students' Motivation, Resilience, and Academic Achievement. *Multidisciplinary Journal of Horseed International University (MJHIU)*, *2*(1), 102-125. https://doi.org/10.59336/7adj0850

Mokkink, L., Terwee, C., & de Vet, H. (2021). Key concepts in clinical epidemiology: Responsiveness, the longitudinal aspect of validity. *J Clin Epidemiol*, *140*, 159-162. https://doi.org/10.1016/j.jclinepi.2021.06.002

Muradoglu, M., Porter, T., Trzesniewski, K., & Cimpian, A. (2024). A Growth Mindset Scale for Young Children (GM-C): Development and validation among children from the United States and South Africa. *PloS one*, *19*(10), e0311205. https://doi.org/10.1371/journal.pone.0311205

Noble, C. (2015). Mindsets, mind sets and mind sense. *Prometheus*, *33*(4), 411–420. https://doi.org/10.1080/08109028.2016.1199379

Rangka, I. B., Hidayah, N., Hanurawan, F., & Eva, N. (2023). A Rasch-based validation of the Indonesian version of the food neophobia scale. *Food Quality and Preference*, *109*(5), 104891. https://doi.org/10.1016/j.foodqual.2023.104891

Rhew, E., Piro, J. S., Goolkasian, P., & Cosentino, P. (2018). The effects of a growth mindset on self-efficacy and motivation. *Cogent Education*, *5*(1), 1–16. https://doi.org/10.1080/2331186X.2018.1492337

Setyorini, E. T., & Khuriyah, K. (2023). The Influence of Teacher Professionalism and Creativity on Student Motivation in Madrasah Ibtidaiyah. *Attadrib: Jurnal Pendidikan Guru Madrasah Ibtidaiyah*, *6*(1), 40-47. https://doi.org/10.54069/attadrib.v6i1.374

Sholihah, M., Pradana, C. R., & Fatikh, M. A. (2025). Application of Client-Centered Theory Using Positive Regard (Acceptance) Techniques to Manage Anxiety in Students Who Are Studying While Working. *Dirasah International Journal of Islamic Studies*, *3*(1), 77-87. https://doi.org/10.59373/drs.v3i1.38

Sigmundsson, H., & Haga, M. (2024). Growth Mindset Scale: Aspects of reliability and validity of a new 8-item scale assessing growth mindset. *New Ideas in Psychology*, *75*(July), 101111. https://doi.org/10.1016/j.newideapsych.2024.101111

Simms, L. J., Zelazny, K., Williams, T. F., & Bernstein, L. (2019). Does the number of response options matter? Psychometric perspectives using personality questionnaire data. *Psychol Assess*, *31*(4), 557-566. https://doi.org/10.1037/pas0000648

Snyder, A. W. (2002). Breaking Mindset. *Mind & Language*, *13*(1), 1-10. https://doi.org/10.1111/1468-0017.00061

Sumintono, B. (2016). *Aplikasi Pemodelan Rasch pada Asesmen Pendidikan: Implementasi Penilaian Formatif (Assessment For Learning). Makalah Dipresentasikan Dalam Kuliah Umum Pada Jurusan Statistika*. Institut Teknologi Sepuluh November. http://eprints.um.edu.my/15876/1/ITS_rasch_model_asesment_for_learning.pdf

Sumintono, B. (2018). Rasch Model Measurements as Tools in Assesment for Learning. *173*, 38–42. https://doi.org/10.2991/icei-17.2018.11

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi Model Rasch untuk Penelitian Ilmu-Ilmu Sosial*. Trim Komunikata Publishing House. Google Scholar

Vestad, L., & Bru, E. (2023). Teachers' support for growth mindset and its links with students' growth mindset, academic engagement, and achievements in lower secondary school. *Social Psychology of Education*, *27*(4), 1431-1454. https://doi.org/10.1007/s11218-023-09859-y

Vyas, R., & Supe, A. (2008). Multiple choice questions: a literature review on the optimal number of options. *Natl Med J India*, *21*(3), 130-133. https://www.ncbi.nlm.nih.gov/pubmed/19004145

Wu, Q., & Tang, M. L. (2016). Non-randomized response model for sensitive survey with noncompliance. *Stat Methods Med Res*, *25*(6), 2827-2839. https://doi.org/10.1177/0962280214533022

YalcİN, S. B., & Yilmaz, E. (2023). A Study on the Predictability of Students' Grit by their Growth Mindset. *Participatory Educational Research*, *10*(5), 184-199. https://doi.org/10.17275/per.23.81.10.5

Yeager, D. S., & Dweck, C. S. (2012). Mindsets That Promote Resilience: When Students Believe That Personal Characteristics Can Be Developed. *Educational Psychologist*, *47*(4), 302-314. https://doi.org/10.1080/00461520.2012.722805

Yilmaz, E. (2022). Development of Mindset Theory Scale (Growth And Fixed Mindset): A Validity and Reliability Study (Turkish Version). *Research on Education and Psychology*, *6*(Special Issue), 1-26. https://doi.org/10.54535/rep.1054235

Zakrajsek, J. A. (2017, 2017). *Promoting Growth Mindset in Middle School Students: An Intervention using Read-Alouds* Proceedings of the Wisconsin Space Conference, https://doi.org/10.17307/wsc.v1i1.232

Zhao, Y. (2006). How to design and interpret a multiple-choice-question test: A probabilistic approach. *International Journal of Engineering Education*, *22*(6), 1281–1286. Google Scholar